(*Reprinted from Nature, Vol.* 171, *p.* 59, *January* 10, 1953)

COMPOUND HELICAL CONFIGURATIONS OF POLYPEPTIDE CHAINS: STRUCTURE OF PROTEINS OF THE α-KERATIN TYPE*

By PROF. LINUS PAULING, For.Mem.R.S., and PROF. ROBERT B. COREY

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena 4

LAST year we described several configurations for polypeptide chains, and suggested that one of them, the α-helix, which has about 3·6 amino-acid residues per turn of the helix, is present not only in synthetic polypeptides but also in proteins of the α-keratin type, and in hæmoglobin and other globular proteins[1]. We pointed out that the dimensions of the unit of structure, as indicated by the X-ray photographs, and the distribution of intensities in the equatorial direction are roughly accounted for by the α-helix. In addition, we mentioned that the structure is supported also by the meridional reflexions observed for muscle fibres and porcupine quill, in particular the reflexions with spacings 1·5 A. (the length per residue along the axis of the α-helix) and multiples of this value; the value of this evidence was emphasized by Perutz[2], who observed the 1·5-A. reflexion for hair, horn, and other α-keratin proteins.

We also pointed out that the X-ray pattern of α-keratin, as described by Astbury and Street[3], has a strong equatorial reflexion at 27 A., which is not accounted for by a structure involving α-helixes in regular parallel orientation, and we suggested that this reflexion is due to a long-range order to be elucidated through further study. Other difficulties for the proposed simple structure of the α-keratin proteins have been emphasized to us by several workers in the field. The most important of these difficulties are a discrepancy between the observed and calculated density, and the failure of the α-helix to explain the 5·2-A. arc on the X-ray photographs as a true meridional reflexion.

We have recently noticed that an α-helix for a polypeptide chain involving repeating sequences of amino-acid residues of different kinds would be expected not to have a straight axis; instead, the

* Contribution No. 1745.

axis of the helix would itself be predicted to pursue a helical course. A protein might consist of a single polypeptide chain with the configuration of a compound helix; three such helixes might well be expected to twist about one another, to form a three-strand rope, and six might be expected to twist about a seventh, to form a seven-strand cable; still more complex structures may also be formed. There is good reason for believing that hair, horn, quill, and other proteins of the α-keratin type consist of seven-strand protein cables in parallel orientation, with single compound α-helixes occupying interstices between them.

The study of simple substances related to proteins has not only provided reliable values of interatomic distances and bond-angles, but has also indicated that the N—H···O hydrogen-bond distance may be expected to vary by about ± 0·12 A. about its average value[4] in compounds of this sort, 2·80 A. Variation in the hydrogen-bond distance might be caused directly by the interaction of side-chains of amino-acid residues with the carbonyl and imino groups of the adjacent amide groups, or indirectly by steric hindrance (especially adjacent to a proline or hydroxyproline residue) or by attraction between side chains. Let us consider an α-helix composed of a polypeptide in which a unit of four amino-acid residues of different sorts is continually repeated. Two of the hydrogen bonds might be longer than the other two, by about 0·2 A. This difference in length would cause a curvature of the axis of the α-helix. If the α-helix has 3·6 residues per turn, the normal to the curved helical axis would be rotated by 0·09 revolution by progressing from one unit to the next unit of four residues along the chain, which corresponds to a complete revolution in about eleven units. The axis of the α-helix would itself accordingly describe a larger helix, with pitch approximately 66 A., the axial length of 44 residues. The radius of the larger helix would be about 1·5 A., and the sense of the larger helix would be the same as that of the α-helix. A compound helix of this sort is represented in Figs. 1 and 2.

Another simple case is that of the compound helix with a repeating unit of seven amino-acid residues. An α-helix with 3·60 residues per turn executes 97·2 per cent of two turns in seven residues, and would accordingly be expected to complete a turn of the larger helix in about thirty-five turns of the α-helix, or 126 residues, corresponding to about 190 A. for the pitch of the helix. However, the prediction of the pitch of this compound helix is rather uncertain; decrease by 1·5 per cent of the number of residues per turn would cause the predicted pitch to be doubled, to the value about 400 A. The radius

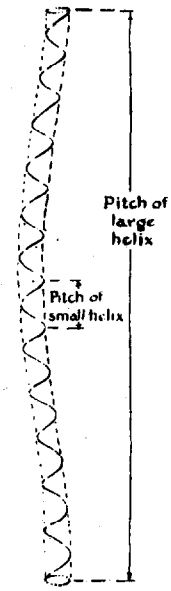

Fig. 1

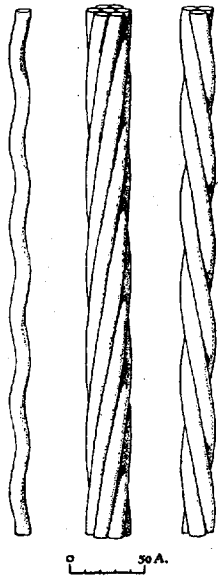

Fig. 2

Fig. 1. A compound helix with pitch of the large superimposed helix equal to 12·5 times the pitch of the small helix. For the compound α-helix the values might be 68 A. and 5·44 A. respectively

Fig. 2. At the left, a compound α-helix with pitch about 67 A. The diameter, shown as about 10 A., includes the volume occupied by side chains as well as the main chains of the protein. Centre, a seven-strand α-cable, with lead of about 400 A. In the proposed structures of proteins of the α-keratin type these cables are packed together, with compound helixes as shown at the left in the interstices. At the right, a three-strand rope, with lead of about 200 A., and with sense opposite to that of the seven-strand cable. In the suggested structure for feather rachis keratin, these ropes are packed together with seven-strand α-cables, in the ratio of two ropes to one α-cable

of the large helix might easily be as great as 10 A., with the variation in hydrogen-bond length mentioned above. The sense of the large helix of the seven-residue compound α-helix is opposite to that of the α-helix itself.

A radius of 6 A. for the large helix would permit three compound helixes to twist about one another to form a three-strand rope (Fig. 2). Such a rope with the sense of twist of the strands opposite to that of the rope would be formed by the seven-residue compound helix, or with the sense of the strands the same as that of the rope by, for example, the fifteen-residue compound helix (with a repeating unit of fifteen residues, comprising nearly four turns of the α-helix).

Six compound helixes with radius of the large helix equal to 10 A. could twist about a central straight α-helix, to form a seven-strand cable. The repeating unit of seven amino-acid residues, comprising nearly two turns of the α-helix, is the simplest one that would give rise to compound helixes suitable to a seven-strand cable; the next simplest is the repeating unit of fourteen (perhaps containing one proline residue). A drawing of the seven-strand α-cable is shown in Fig. 2. We suggest that the symbol AB_6 be used for this cable.

The seven-strand α-cable is about 30 A. in diameter. A fibre containing these cables in parallel orientation would have a hexagonal unit of structure or pseudo-unit with a equal approximately to 30 A. The observed equatorial reflexion with spacing 27 A. could then be explained as the 10·0 reflexion corresponding to this unit. The X-ray patterns of hair, horn, porcupine quill, and other α-keratin proteins are reasonably well accounted for by such a unit, with $a = 32{\cdot}4$ A. (for porcupine quill a multiple of this unit, indicating further superstructure, is needed to explain weak reflexions with larger spacing).

A plan of the packing of the seven-strand α-cable is shown in Fig. 3. It is seen that there is room enough at the positions $\frac{1}{3}\,\frac{2}{3}$ and $\frac{2}{3}\,\frac{1}{3}$ for other polypeptide chains to be introduced. If the adjacent seven-strand cables are staggered in azimuthal orientation, at a given level, α-helixes may be introduced in the positions $\frac{1}{3}\,\frac{2}{3}$ and $\frac{2}{3}\,\frac{1}{3}$. Because of the rotation of the cables, for which the value of the

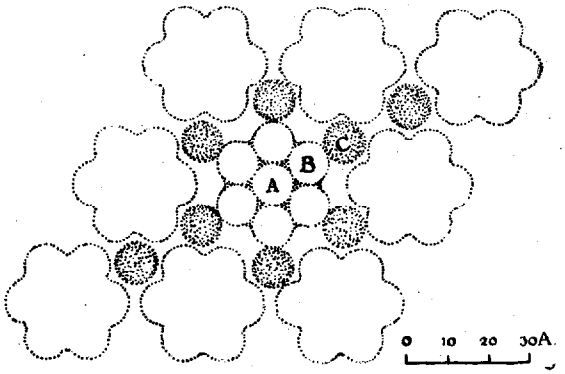

Fig. 3. A cross-section of the α-keratin structure, showing the α-cables AB_6 and the interstitial compound helixes C. The orientation of the cross-section of the cable changes with co-ordinate along the fibre axis. The central cable is shown in the most unfavourable orientation for the interstitial α-helixes. The protein chains are not so nearly circular in cross-section as indicated in the drawing, and space is filled more effectively than is indicated

lead (vertical distance between a point on one strand and an equivalent point on the same strand after one turn about the cable) is about 400 A., and that of the pitch (vertical distance from one strand to an adjacent strand) is about 66 A., the inserted α-helixes must be compound helixes, with radius of the large helix about 1·5 A., and with pitch equal to the pitch of the cable, about 66 A. The sense of these compound helixes must be opposite to that of the cables, in order that they may fit between the strands of an adjacent cable. The four-residue compound helix, which has the proper sense and approximately the correct pitch, is accordingly satisfactory. We suggest the symbol *C* for this compound helix.

The volume of a portion of the hexagonal unit with $a = 32{\cdot}4$ A., and with height 1·50 A., the axial length per residue, is 136·3 A.^{3}. The proposed structure involves nine amino-acid residues in this portion of the unit. With the reported range 107–118 for the average residue weight, the density of the protein is calculated to be 1·17–1·30 gm.cm.$^{-3}$. The indicated disagreement with the experimental value, about 1·32 gm.cm.$^{-3}$, may be due to the presence of a few per cent of water of hydration in the fibrous proteins. The much larger discrepancy between calculated and observed density that results from assuming the centre of the strong but broad equatorial reflexion, at 9·8 A., to correspond to the spacing of the reflexion 10·0 for the small hexagonal unit containing one α-helix is eliminated by the new indexing, which assigns the strong reflexion with centre at 9·8 A. to the overlapping of 21·0 at 10·6 A., and 33·0 at 9·3 A.

The strong 5·2-A. meridional arc, characteristic of the α-keratin proteins and previously given only rather unsatisfactory explanation in terms of the α-helix, is now explained in a straightforward manner as resulting from the co-operation of the seven residues, in a repeat involving approximately two turns of the α-helix, in the *B* compound helixes of the AB_6 cable. With 1·50 A. as residue length of the α-helix, the unit of seven residues has length 10·5 A. The component of this distance in the direction of the fibre axis, assuming the cable to have lead 400 A. and radius 10 A., is calculated from the angle of inclination, 9·0°, to be 10·36 A. A weak meridional reflexion would be expected to occur with this spacing; the second order of this reflexion, with spacing 5·18 A., should be strong, because it involves reinforcement rather than extinction by the two turns of the α-helix in the repeating unit of seven residues. The 5·2-A. meridional arc, for which Astbury and MacArthur have recently reported spacings of about 5·15 A. or 5·18 A., is accordingly well explained by the proposed structure.

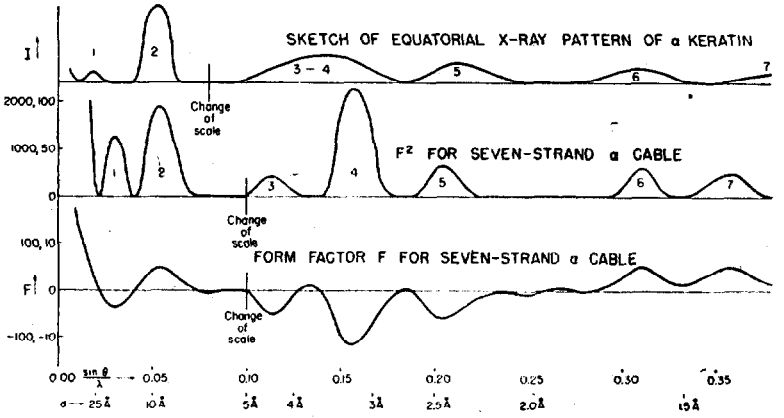

Fig. 4. Comparison of observed and calculated equatorial intensities on X-ray photographs of α-keratin. At the top there is a sketch of the observed intensity of reflexion by hair and horn. The decrease in intensity with increase in angle is more rapid than indicated. At the bottom, there is given the calculated form factor for the AB_6 cable, and in the centre the square of this form factor, which may be compared with the observed intensity curve

Strong support of the proposed structure is also given by many other features of the X-ray patterns of various proteins of the α-keratin type. One comparison of observed and calculated features is given in Fig. 4. At the top of this figure there is a sketch of the equatorial X-ray pattern of α-keratin. Astbury and Street[3] described the α-keratin pattern as having a strong reflexion at 27 A., a very strong and broad reflexion at about 9·8 A., and a vague region of darkening around 3·5 A. Examination of heavily exposed photographs shows that the last of these features is, in fact, a broad maximum, extending from about 5 A. to about 3·2 A. (This band may be due in part to the very strong 4·7-A. reflexion of β-keratin; it is known that specimens of α-keratin usually contain some β-keratin, and some of our photographs show a rather strong 4·7-A. ring.) There is then a minimum at about 2·8 A., a maximum at about 2·4 A., a minimum at about 1·8 A., a maximum at about 1·6 A., a minimum at 1·4 A., and a further maximum. The form factor F for the α-cable shown at the bottom of the figure is calculated by multiplying the form factor for the α-helix, including the β-carbon atom, as reported previously[1], by the factor $1 + 6J_0(4\pi\rho \sin \theta/\lambda)$, in which J_0 is the Bessel function of order 0, and ρ, the radius of the B helixes of the cable, is placed equal to 10 A. Above the curve for F there is shown the curve for F^2. It is seen that there is striking, although not complete, agreement between this curve and the sketch of the equatorial X-ray pattern. Inasmuch as the contribution of the two interstitial compound α-helixes and of the side chains, which together constitute just

50 per cent of the scattering matter in the fibres, have not been taken into consideration, the agreement must be considered to be excellent. It may be pointed out that, in the calculation of the form factor, no arbitrary parameters have been involved except the radius of the large helix; and that the value 10 A. for this quantity is required, to within a few per cent, by the dimensions of the unit of structure, if it is assumed that the three different kinds of polypeptide chains in the α-keratin proteins (the core of the cable, the six surrounding strands, and the interstitial compound helixes) have approximately the same average amino-acid residue weight.

The X-ray diffraction photographs of feather rachis keratin, which have previously been interpreted as involving β-keratin pleated sheets and α-keratin helixes in molecular distribution, may show a superimposed β-keratin pattern and α-keratin pattern. If this is correct, the α-keratin is different in nature from hair and horn α-keratin, probably consisting of AB_6 cables with D_3 ropes (three α-helixes coiled about one another) in the interstices. This structure accounts in a striking way for the characteristic features of the X-ray photographs.

The proposed structure of α-keratin involves three kinds of α-helixes, which presumably differ in amino-acid composition. We propose to examine these materials by fractionation. It is interesting that Goddard and Michælis have reported that wool put into solution by reduction with thioglycollic acid or other reducing agent and treated with iodoacetic acid to form a carboxymethyl derivative can be converted into fractions with different chemical composition by ammonium sulphate precipitation[5]. We suggest that a solution of wool might be fractionated into the AB_6 cables and the protein keratin-C, and that the AB_6 cables might be further separated into keratin-A and keratin-B.

We think it not unlikely that actomyosin has the structure shown in Fig. 3, and that its fractionation into myosin and actin is a separation of the seven-strand cables AB_6 (myosin) from the single α-helixes C (actin). The amounts of myosin and actin obtained by the fractionation are approximately in the predicted ratio 7 : 2, and, as predicted, the 5·2-A. meridional X-ray reflexion is observed for actomyosin and myosin but not for actin.

We are indebted to Prof. W. T. Astbury, Dr. I. MacArthur, Mr. F. H. C. Crick, and other workers in the field of the structure of proteins for having discussed with us the question of the structure of α-keratin, and especially for having emphasized the necessity for refinement of our suggestions. The detailed description of compound α-helixes and

aggregates of them, and discussion of fibrous proteins which they compose, will be presented in later papers, to be published in the *Proceedings of the National Academy of Sciences* of the United States of America.
This investigation was supported in part by a research grant from the National Institutes of Health, Public Health Service, and by a contract (Nonr-220(05)) with the Office of Naval Research. [Oct. 14.

[1] Pauling, L., and Corey, R. B., *J. Amer. Chem. Soc.*, **72**, 5349 (1950). Pauling, L. Corey, R. B., and Branson, H. R., *Proc. U.S. Nat. Acad. Sci.*, **37**, 205 (1951). Pauling, L., and Corey, R. B., *ibid.*, **37**, 235, 241, 251, 256, 261, 272, 282, 729 (1951); *ibid.*, **38**, 86 (1952); *Nature*, **168**, 550 (1951); **169**, 494, 920 (1952).

[2] Perutz, M. F., *Nature*, **167**, 1053 (1951).

[3] Astbury, W. T., and Street, A., *Phil. Trans. Roy. Soc.*, A, **230**, 75 (1951).

[4] Donohue, J., *J. Phys. Chem.*, **56**, 502 (1952). Corey, R. B., and Pauling, L., *Proc. Roy. Soc.* (in the press).

[5] Goddard, D. R., and Michaelis, L., *J. Biol. Chem.*, **106**, 605 (1934); **112**, 361 (1935).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.